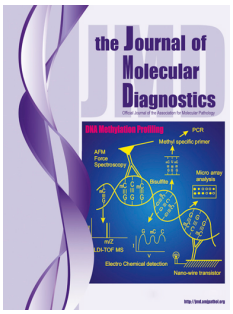
Design and development of a fully synthetic MLPA-based probe mix for detection of copy number alterations in prostate cancer formalin-fixed, paraffin-embedded tissue samples

Journal Pre-proof



Design and development of a fully synthetic MLPA-based probe mix for detection of copy number alterations in prostate cancer formalin-fixed, paraffin-embedded tissue samples

Walead Ebrahimizadeh, Karl-Philippe Guérard, Shaghayegh Rouzbeh, Yogesh M. Bramhecha, Eleonora Scarlata, Fadi Brimo, Palak G. Patel, Tamara Jamaspishvili, Armen G. Aprikian, David Berman, M S John Bartlett, Simone Chevalier, Jacques Lapointe

PII: S1525-1578(20)30426-8

DOI: https://doi.org/10.1016/j.jmoldx.2020.07.003

Reference: JMDI 961

To appear in: *The Journal of Molecular Diagnostics*

Received Date: 14 February 2020

Revised Date: 24 June 2020

Accepted Date: 15 July 2020

Please cite this article as: Ebrahimizadeh W, Guérard K-P, Rouzbeh S, Bramhecha YM, Scarlata E, Brimo F, Patel PG, Jamaspishvili T, Aprikian AG, Berman D, Bartlett J, Chevalier S, Lapointe J, Design and development of a fully synthetic MLPA-based probe mix for detection of copy number alterations in prostate cancer formalin-fixed, paraffin-embedded tissue samples, *The Journal of Molecular Diagnostics* (2020), doi: https://doi.org/10.1016/j.jmoldx.2020.07.003.

Design and development of a fully synthetic MLPA-based probe mix for detection of copy number alterations in prostate cancer formalin-fixed, paraffin-embedded tissue samples

Walead Ebrahimizadeh[1¶], Karl-Philippe Guérard[1], Shaghayegh Rouzbeh[1‡], Yogesh M Bramhecha[1¶], Eleonora Scarlata[1], Fadi Brimo[2], Palak G Patel[3], Tamara Jamaspishvili[3], Armen G Aprikian[1], David Berman[3], John M S Bartlett[4,5,6], Simone Chevalier[1] and Jacques Lapointe[1].

[1]Department of Surgery, Division of Urology, [2]Department of Pathology; McGill University and the Research Institute of the McGill University Health Centre (RI MUHC), Montreal, Quebec, Canada.
[3]Department of Pathology, Queen's University, Kingston, Ontario, Canada.
[4] Diagnostic Development, Ontario Institute for Cancer Research, Toronto, Ontario, Canada.
[5] Laboratory Medicine & Pathobiology, University of Toronto, Toronto, Ontario, Canada
[6] Edinburgh Cancer Research Centre, University of Edinburgh, Edinburgh, United Kingdom.

[¶] current affiliation of W.E., IMV Inc. Canada; of S.R., Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM)

Running title: Prostate cancer specific MLPA probe mix

Funding: This work was proudly funded by the Movember Foundation through Prostate Cancer Canada (grant #T2014.01) to D.B., J.M.S.B, and J.L and was carried out as a part of personalized risk stratification for patients with early prostate cancer (PRONTO). W.E. received a studentship and doctoral training award from the Research Institute of McGill University Health Centre and Fonds de Recherche du Québec-Santé (FRQS), respectively.

Corresponding Author:
Jacques Lapointe, MD. Ph.D.,
Department of Surgery, Division of Urology,
Research Institute of the McGill University Health Centre,
EM2.2212, 1001 Boulevard Décarie, Montréal, QC H4A 3J1, Canada
Telephone: 514-934-1934 ext. 44638; Fax: 514-933-2691
Email: jacques.lapointe@mcgill.ca

Disclosure/Conflict of Interest

The authors declare that they have no conflict of interest.

Abstract

DNA copy number alterations (CNAs) are promising biomarkers to predict prostate cancer (PCa) outcome. However, fluorescence *in situ* hybridization (FISH) cannot assess complex CNA signatures due to low multiplexing capabilities. Multiplex ligation-dependent probe amplification (MLPA) can detect multiple CNAs in a single PCR assay, but PCa-specific probe mixes available commercially are lacking. Synthetic MLPA probes were designed to target 10 CNAs relevant to PCa: 5q15-21.1 (*CHD1*), 6q15 (*MAP3K7)*, 8p21.2 (*NKX3-1)*, 8q24.21 (*MYC)*, 10q23.31 (*PTEN)*, 12p13.1 (*CDKN1B)*, 13q14.2 *(RB1)*, 16p13.3 *(PDPK1)*, 16q23.1 (*GABARAPL2)*, and 17p13.1 *(TP53)* with 9 control probes. In cell lines, CNAs were detected when the percentage of cancer genome was as low as 30%. Compared to FISH in radical prostatectomy (RP) formalin-fixed paraffin-embedded (FFPE) samples (n=18: 15 cancers, 3 matched benign), the MLPA assay showed median sensitivity and specificity of 80% and 93%, respectively, across all CNAs assessed. In the validation set (n=40: 20 tumors sampled in two areas), the respective sensitivity and specificity of MLPA compared advantageously to FISH and digital droplet TaqMan PCR (ddPCR) when assessing *PTEN* deletion (FISH: 85% and 100%, ddPCR: 100% and 83%) and *PDPK1* gain (FISH: 100% and 92%, ddPCR: 93% and 100%). In conclusion, this new PCa probe mix accurately identifies CNAs by MLPA across multiple genes using low quality and quantities (50 ng) of DNA extracted from clinical FFPE samples.

Introduction

Prostate cancer (PCa) is the most commonly diagnosed visceral malignancy in North American men and a leading cause of cancer-related mortality [1, 2]. PCa has a heterogeneous clinical outcome ranging from an indolent disease to a deadly metastatic cancer. Standard treatments for localized PCa consist of radical prostatectomy (RP) or radiotherapy, which are often associated with significant adverse side effects. Active surveillance, a close monitoring of the disease with intent for therapeutic interventions at first sign of progression, has become a viable option for patients with low-risk PCa [3]. The key challenge in the management of early PCa is to distinguish between indolent disease and clinically significant tumors. Current prognostic parameters such as pre-operative PSA levels, clinical staging and Gleason grading of biopsy specimens cannot accurately predict individual clinical outcome [4], which leads to overtreatment of clinically insignificant diseases and under-treatment of aggressive cancers with metastatic potential as a result of incomplete risk assessment at diagnosis [5]. Better prognostic indicators would allow immediate treatment of patients with potentially life-threatening cancer and active surveillance of patients who would not benefit from immediate or aggressive treatments. Thus far, genomic profiling has shed some light on the molecular heterogeneity of PCa and has identified potential prognostic markers that can improve risk stratification and provide better assessment of disease status. Loss of expression of tumor suppressor protein such as PTEN can be assessed on tissue section by a standard immunohistochemistry assay and has been associated with poor clinical outcome [6]. To assess simultaneously the expression of multiple biomarkers and improve patient stratification, RNA-based assays such as Prolaris [7], Oncotype Dx Genomic Prostate Score [8], and Decipher genomic classifier [9] have been developed. However, the development of multiplex DNA-based assays has lagged behind RNA-based markers even though DNA holds several advantages such as its resistance to degradation and its stability in various physiological and environmental conditions.

DNA copy number alterations (CNAs) are genomic alterations consisting of deletion or gain of genomic DNA segments ranging from one kilobase to several megabases that may influence the function of several genes and regulatory elements [10] and impact disease progression [11, 12]. In PCa, CNAs are more common than single nucleotide alterations [13, 14] and their frequency increases in advanced cancers [13-16]. Gain of oncogenes such as *MYC* and loss of tumor suppressors such as *PTEN* and *TP53* are examples of CNAs that are associated with disease progression and may serve as prognostic biomarkers in PCa [17-23]. Moreover, assessing the CNA status of *PTEN* would improve the outcome prediction of its expression measured by immunohistochemistry [6]. Furthermore, studies suggest that the assessment of a combination of multiple CNAs in primary tumors improves patient risk stratification [18, 24-26]. The gold standard method for detection of CNAs in clinical tissue samples is fluorescence *in situ* hybridization (FISH) [24]. However, FISH sensitivity and resolution are probe dependent and implementation for high throughput analysis has been limited by the cost of reagents, labor and the small number of CNAs that can be assessed in a single assay [27]. Given the small amount of tissue obtained from prostate biopsies, there is an unmet need for new multiplex assays to apply CNA biomarker signatures in the clinical setting. However, the development of such assays based on extracted nucleic acids must confront the challenge of prostate tumor tissues heterogeneity that often include benign glandular and stromal cells alongside cancer cells. Thus, genomic DNA extracted from such samples consists of cancer cell DNA diluted to various degrees with benign cell DNA.

Multiplex ligation-dependent probe amplification (MLPA) is a multiplex PCR-based CNA detection method [27]. MLPA probes consist of two “half-probes”, flanked by universal primer binding sites, which hybridize next to each other to a specific DNA sequence target (CNA region). Upon ligation of the two half probes, which can only occur if the DNA target sequence is present in the sample, the complete probe is PCR amplified using fluorescently labeled primers. By addition of a stuffer sequence at the design step, each MLPA probe is given a unique

size and fluorescent PCR products corresponding to target sequences can be separated and quantified by capillary electrophoresis. MLPA can assess up to 50 different loci in a single reaction with a resolution of 30-60 bp and is compatible with low amounts (50 ng) of fragmented DNA extracted from FFPE biopsy samples. This assay is low-cost, and results can be generated within two days using a multi-well format suitable for high throughput analysis. MLPA has been used to simultaneously detect multiple CNAs in a variety of cancers [28, 29]. However, to our knowledge there is currently no PCa-specific probe mix available to assess multiple PCa relevant CNAs. In this study, we report the design, validation and application of a PCa specific MLPA probe mix that can be used on FFPE samples to assess ten CNAs relevant to this disease.

Materials and Methods

Cell lines

Human prostate cancer cell lines PC-3, DU145, and LNCaP were purchased from American Type Culture Collection (ATCC, Gaithersburg, MD) and the LAPC4 cell line was obtained from Dr. Robert Reiter, Department of Urology, University of California, Los Angeles. All cell lines were cultured in Hyclone RPMI 1640 media (GE Life Sciences, Chicago, IL) supplemented with 10% fetal bovine serum (Wisent Bioproducts, Saint-Jean-Baptiste, Canada), Gibco 1% penicillin-streptomycin, and 2 mM L-Glutamine (Thermo Fisher Scientific, Bedford, MA).

Prostate tissue samples

This study was approved by the Research Ethics Board of McGill University Health Centre (Quebec, Canada, BDM-10-115) and amended to include samples from Kingston General Hospital collected with the approval of the Queen's University Research Ethics Board (Ontario,

Canada). All RP FFPE specimens were histologically reviewed centrally by pathologists to identify the cancer area and assign the final Gleason grade, according to the latest International Society of Urological Pathology/World Health Organization recommendations [30]. The samples from Queen's University were used to develop the method and hereafter will be referred to as the test sample set (n=18: 15 cancers and 3 matched benign areas, Table 1). Tissue cores of 0.6 mm were harvested from the same RP FFPE block to extract DNA for MLPA (3 cores) and to build a tissue microarray (TMA) used for FISH analyses (3 cores). For the method validation, an independent set of RP FFPE tissue specimens (n=40: 20 tumors sampled in two areas, Table 1) collected at the McGill University Health Centre were cored (3 x 0.6 mm, 2-3mm in length) for DNA extraction in two tumor areas representing the highest (A sample) and lowest (B sample) Gleason grade patterns identified by the pathologists. When only one Gleason grade pattern was represented, the two sampled areas were assigned to A or B randomly. The FISH on the validation samples was then performed on tissue sections of the same blocks used for DNA extraction.

DNA extraction

DNA from cell lines was extracted using Qiagen DNeasy Blood & Tissue Kit (Qiagen Inc., Toronto, Canada). Two million cells grown in monolayers were detached using Gibco trypsin and washed twice with PBS before DNA extraction according to Qiagen manufacturer instructions. For FFPE tissue samples, the DNA was extracted from three 0.6 mm cores with a modified protocol of the AllPrep DNA/RNA FFPE Kit (Qiagen) as previously described [31].

Normal reference DNA samples

Commercially available normal female genomic DNA (Promega, Madison, WI), hereafter referred to as fresh DNA, DNA extracted from FFPE pathology punches of normal kidney

(McGill University Health Centre) and normal breast lymph nodes (Ontario Tumor Bank, Ontario Institute for Cancer Research) were used as reference samples for MLPA experiments. To assess the CNA detection limit, commercially available normal male genomic DNA (Promega) was used to dilute PC-3 genomic DNA.

MLPA probe selection, design, and synthesis

Due to technical limitation in the synthesis of long oligonucleotides with high coupling efficiency, synthetic MLPA probe mixes generally include 20-30 probes [32-34]. Thus, to build a PCa specific MLPA probe mix and considering the number of synthetic probes that can be included, we performed a literature review and selected ten relevant genes (loci) that undergo CNA in PCa and are known to be associated with clinical outcome or have the potential to improve patient risk stratification. These include the known oncogene *MYC* (8q24.21) [17, 21, 23], the tumor suppressors *PTEN* (10q23.31) [18, 20, 22, 35], *TP53* (17p13.1) [19, 21, 36], *CDKN1B* (12p13.1) [37, 38], and *RB1* (13q14.2) [15, 39], genes in loci associated with metastasis such as *GABARAPL2* (16q23.1) [15, 25] and *PDPK1* (16p13.3) [40, 41], and genes associated with the previously described molecular PCa subtypes [15] *CHD1* (5q15-q21.1) [15, 42, 43], *MAP3K7* (6q15) [15, 42, 44] and *NKX3-1* (8p21.2) [15] (Table 2).

For normalization and data analysis, MLPA also requires internal reference probes targeting loci that are less likely to undergo CNA. Most synthetic MLPA probe mixes use a commercially available reference mix named P200 from MRC-Holland (MRC Holland, Amsterdam, Netherlands) for this purpose. The P200 reference mix is unsuitable for PCa since it contains probes that target loci commonly deleted in PCa, such as 21q22 (*TMPRSS2-ERG* fusion) [45-47], 10q22, 5q31 and 16q24 [15, 16]. Therefore, three publicly available PCa CNA datasets [15, 16, 26] were analyzed to select genes least affected by CNA (≤5%) and were used to design nine new reference probes (Table 2).

MLPA probes were designed using MLPA Designer 7.91 software (PREMIER Biosoft, Palo Alto, CA), according to guidelines provided by MRC-Holland. All genomic sequences were obtained from the NCBI Reference Sequences (RefSeq) database. Twenty probes were designed for each exon of the selected genes. The GC content was chosen to be between 40% and 65% and the Tm was 72 ± 5ºC for all probes. The assay was designed to suit DNA fragments shorter than 200 bp typically extracted from FFPE samples [31] and therefore the combined hybridization sequences of the 5' half-probe and 3' half-probe of all probes were kept under 200 nucleotides.

The top-ranking probes based on GC content, Tm, hairpin ΔG and probes terminal nucleotide sequences were assessed for probe-probe interactions. Probes showing interactions stronger than -3 ΔG were replaced with the next ranking exon probes available. In instances that exon probes could not be selected, intron probes were designed.

Two probes for each CNA gene targeting two different genomic regions and one probe per reference gene were selected in the final probe mix. Stuffer sequences as previously reported by Zhi and Hatchwell [48], were added between the hybridization sequence and the universal primer binding sequences of each half-probe to give a unique length to the final probe with at least three nucleotide length differences between consecutive probes. The lengths of the complete MLPA probes (ligated 5' and 3' half-probes including stuffer sequences) were kept under 260 nucleotides, since longer synthetic probes did not yield reproducible results (data not shown). Probe specificity was confirmed by BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start, accessed on October 9th, 2017). The presence of known PCa mutations and SNPs within the probe target sequence was ruled out using Ensembl Variation database [49] and PCa SNP data downloaded from cBioPortal website, respectively [50, 51].

All probes were synthesized and purified using standard desalting by Integrated DNA Technologies (IDT, Coralville, IA) and received at 4 nmole scale. The 3' half-probes were 5' phosphorylated. Probes specifications including the coordinates of their target sequences

according to the 2013 UCSC Genome Browser assembly (hg38) are available in Table 2. All probes were dissolved and diluted in TE_1 buffer (10mM Tris-HCl, 1 mM EDTA; pH 8.0). A synthetic probe mix was constructed by adding 0.8 µl of 1 µM concentration of each half-probe and 1:25 dilution of the CF004-A1 control fragments (Q+D-control fragments, Lot: A1-0715, MRC-Holland), allowing assessment of DNA quantity and quality of MLPA reactions. The final volume of the probe mix was then adjusted to 600 µl using TE_1 buffer. The newly designed probe mix was named "PCa probe mix".

MLPA reaction and analysis

The MLPA reaction was performed according to the MLPA General Protocol (One-Tube) using MLPA kit EK1-FAM (MRC-Holland). Components of the kit and buffers are available in MLPA General Protocol (MDP version-007; Issued on 01 March 2019) at the manufacturer website. Unless otherwise specified, each experiment included duplicate reactions for each sample and four repeats of each reference sample.

Briefly, for each reaction, 50 ng of DNA in 5 µl of $TE_{0.1}$ buffer (10mM Tris-HCl; 0.1 mM EDTA; pH 8.2) was used. After initial denaturation at 98^{o}C for 5 min, the hybridization mix (1.5 µl of PCa probe mix and 1.5 µl of SALSA MLPA buffer per reaction; as per manufacturer instructions) was added at room temperature. After denaturation at 95^{o}C for 1 min, the hybridization was performed at 60^{o}C for 16 h. The ligation mix (25 µl of ddH2O, 3 µl of Ligase buffer A, 3 µl of Ligase buffer B and 1 µl of SALSA Ligase-65 per reaction) were then added at 54^{o}C followed by incubation at 54^{o}C for 15 minutes and then 98^{o}C for 5 minutes. The polymerase mix (7.5 µl of ddH2O, 2 µl of PCR primer mix and 0.5 µl of SALSA polymerase per reaction) was added at room temperature and the PCR reaction was carried out for 35 cycles of 95^{o}C for 30 sec, 60^{o}C for 30 seconds and 72^{o}C for 1 minute followed by a final extension of 20 minutes at 72^{o}C. For all reactions, a Bio-Rad MyCycler thermal cycler and Bio-Rad PCR plates

(Cat. No. 2239441) and caps (Cat. No. TCS0803) were used. For detection of the error-rate of the assay, another thermal cycler (T100 Bio-Rad, Hercules, CA), different PCR plates (Cat. No. AB0600, Thermo Fisher Scientific), caps (Cat. No. AB0265, Thermo Fisher Scientific), mineral PCR oil (Vapor-Lock, Qiagen) and a different reference probe mix (P200 – B1, Lot: B1-1215, MRC-Holland) were also used.

Capillary electrophoresis was performed by the Genomics platform of the Institute for Research in Immunology and Cancer (IRIC), Université de Montréal using GeneScan 500 LIZ dye Size Standard molecular weight marker (Thermo Fisher Scientific) and ABI 3730 DNA analyzer equipped with G5 filter set.

Coffalyser software (Version 140721.1958, MRC-Holland) was used for fragment and comparative analyses of CNAs, using default settings unless otherwise indicated. The manual probe recognition method was applied. Intra sample normalization was performed by calculating the median value of the test probe over each of the reference probes and inter sample normalization was performed by calculating the average value of the normalized probe signal in test sample over each of the reference samples. The classic and the P.I.N.P.2 analysis protocol were used for CNA analysis of cell lines and FFPE samples, respectively.

Four replicates of fresh healthy female genome were used as a reference population in the analysis of cell lines and four replicates of each reference sample (fresh female genome, DNA extracted from normal FFPE kidney and breast lymph node reference samples) were used as reference population for analysis of FFPE samples. For testing further analysis approaches, probe ratios, standard deviations and CNA calls based on cutoff points (probe ratios below 0.65 are considered deletion and above 1.3 are considered as gain) and CNA calls based on the 95% confidence intervals of probes (as described below) were exported to Microsoft Excel 2017.

For the 95% confidence interval approach, gain or deletion calls were assigned for each probe according to the Coffalyser software if this value in the test samples was respectively above or

below those observed in reference samples. If the 95% confidence intervals in the sample and the reference overlapped, normal copy number calls were assigned. In FFPE tissue samples, only MLPA calls for gain of *MYC* and *PDPK1* as well as deletion of the remaining eight targeted genes were considered as CNAs in the analyses.

Reactions showing high standard deviation (>10%) in more than four probes were considered to fail the quality control (QC) and were repeated (if sufficient DNA was available) or removed from the analysis.

For comparison with existing datasets, the probe ratios resulting from the MLPA analysis of the four PCa cell lines were log2 transformed to compare with two CNA datasets generated by array-CGH (Array comparative genomic hybridization): Zhao et al [52] and the cancer cell line encyclopedia (CCLE) [53]. Pearson correlation coefficients between all data points were calculated.

Fluorescence *in situ* hybridization (FISH)

The *TP53* CNA (17p13.1) was assessed using TP53 / CEP 17 FISH probe kit from Abbott Molecular (Des Plaines, IL).

The following BAC clones mapping to the remaining nine target genes were labeled with Spectrum Orange dUTP (Enzo Life Science, Farmingdale, NY) using Nick Translation kit (Abbot Molecular) and used for FISH: RP11-813D1 (5q15-5q21.1; *CHD1*), RP1-154G14 (6q15; *MAP3K7*), RP11-325C22 (8p21.2; *NKX3-1*), RP11-440N18 (8q24.21; *MYC*), CTD-2557P6 (10q23.31; *PTEN*), RP11-180M15 (12p13.2-p13.1; *CDKN1B*), RP11-893E5 (13q14.2; *RB1*), RP11-16C11 (16q23.1; *GABARAPL2*), and RP11-20I23 (16p13.3, *PDPK1*).

The following control probes for each chromosome were used: Centromere enumeration probes labeled with Spectrum Green CEP6, CEP8, CEP10, and CEP12 (all from Abbot Molecular), green 13qtr subtelomere probe (Cytocell, Cambridge, United Kingdom) as well as RP11-530D2

(5p12) and pHuR-195 (16qh) labeled with Spectrum Green dUTP. The *PTEN* CNA in cell lines was detected using the XT PTEN/GRID1 probe (Metasystems, Newton, MA).

Specificity of FISH probes was confirmed on normal metaphase chromosome preparations (Molecular Genetics Laboratory, McGill University Health Centre). Metaphase spreads of PC-3 and LAPC4 cells were prepared for FISH hybridization. Briefly, cells were grown in standard culture medium described above supplemented with 1 µg/ml of Invitrogen Colcemid (Thermo Fisher Scientific) for three hours and then washed with PBS and detached using Gibco Trypsin. After a centrifugation step at 1000 x g for 5 minutes, 5 million cells were resuspended in 2 ml of 0.56% KCl solution and incubated at room temperature for 5 minutes. Cells were re-centrifuged (1000 x g, 5 minutes) and fixed in 0.5 ml of methanol:glacial acetic acid (3:1) solution before being spread on slides and air dried. Slides were then pretreated by incubation in 70% formamide and saline-sodium citrate (SSC) solution (0.3 M sodium chloride and 30 mM trisodium citrate, pH 7.0) at 75°C for 5 minutes followed by dehydration in ethanol series of 70%, 85% and 100% for one minute each. The test and control FISH probes along with the target DNA were codenatured at 73°C for 6 minutes and left to hybridize overnight at 37°C using the ThermoBrite System (Abbott Molecular). Post-hybridization washes were performed in 2X SSC and 0.3% NP-40/0.4X SSC at 73°C for 2 minutes and 1 minute, respectively, followed by a 30-second incubation at room temperature in 2X SSC. Dual-color tissue FISH was performed on 5µm sections of TMAs or whole tissue blocks as we previously reported [40].

FISH data analysis

To evaluate copy number status, fluorescent signals were counted in 100 non-overlapping interphase nuclei for each case (as identified on corresponding H&E staining for tissues) counterstained with ProLong Diamond antifade reagent with DAPI (Thermo Fisher Scientific), to delineate nuclei. Deletion was defined as ≥15% of nuclei containing one or no test locus signal

and by the presence of two control signals as we previously reported [40]. A gene was considered homozygously deleted if ≥15% of nuclei had no test locus signal and two control signals. A gain was defined as present at a threshold of ≥15% of nuclei containing three or more test locus signals and by the presence of two control signals. Images were acquired with an Olympus IX-81 inverted microscope at 96X magnification, using Image-Pro Plus 7.0 software (Media Cybernetics, Rockville, MD).

Digital droplet TaqMan PCR

Digital Droplet PCR for CNA analysis of *PTEN* and *PDPK1* was carried out using TaqMan probes targeting the same regions assessed by the MLPA probes. For the *PTEN* gene, probe (5'-/6-FAM/AGAAAGCTTACAGTTGGGCCCTGT/Iowa Black/-3') and primers (forward: 5'-TCTGTCGCCATGGCTTATTC-3', reverse: 5'-CACCAAGACCCTGTCTCAAA-3') were designed to target exon 9 and for the *PDPK1* gene, probe (5'/6-FAM/CGTGTACGGAGTTCCACTTTCCATGA/Iowa Black/-3') and primers (forward: 5'-AGCAGCTTACATGTCTGAAGTTA -3', reverse: 5'-TGTTCAAGAGGAGCTACAAAGG-3') were designed to target the intron 10 region. For inter-sample normalization, a probe targeting *AGO1* (1p34.3) (5'-HEX/ CAAGTCCAGTGACCACACTCCCAG/ Iowa Black-3', forward primer: 5'-GAAGATGATGCTCAACATTGATGG-3', reverse primer: 5'-AGAGCTGGGAGGGATGAG-3') was used. Test and control probes were respectively labeled with 6-FAM (Fluorescein amidite) and HEX (Hexachloro-Fluorescein) dyes (IDT). DNA extracted from normal kidney (FFPE tissue) was used as the reference sample. Fresh female genome and PC-3 cell line DNA were used as controls.

The assay was carried out by the Genomics platform of IRIC, Université de Montréal according to Bio-Rad TaqMan qPCR instructions (Bio-Rad) with some modifications as described below. Amplification was performed in a 20 µl multiplex reaction containing 6 ng of purified DNA,

500 nM of primers and 250 nM of probes, 2X ddPCR Supermix for probes (no UTP) and 5 unit/reaction XhoI enzyme (New England Biolabs, Ipswich, MA). Samples were subjected to droplet generation by an automated droplet generator. End-point PCR was performed with cycling steps as follows: initially an enzyme activation at 95°C for 10 minutes followed by 50 cycles of denaturation, annealing and extension (each cycle at 95°C for 30 s; 58°C for 1 min; 72°C for 30sec; 2.5°C/sec ramp rate) and finally enzyme deactivation at 98° C for 10 min. Droplets were read on droplet reader QX200 (Bio-Rad) and data were analyzed using QuantaSoft Software (Bio-Rad) which determines the numbers of positive and negative droplets for each fluorophore in each sample. The fraction of positive droplets was then fitted to a Poisson distribution in QuantaSoft Software (Bio-Rad) to determine the absolute copy number in units of copies/µl.

Average ratios of duplicate reactions were used for normalization. Intra sample normalization was done by calculating the ratio of the test probe over *AGO1* reference probe and inter sample normalization was done by calculating the normalized probe ratio in the test sample over the normalized probe ratio obtained from the kidney reference sample. Cutoff for gains (>1.2) and deletion (<0.8) corresponded to two standard deviations above and below the average ratio obtained from samples (fresh female DNA and tumor DNA from FFPE blocks) with no gains and deletions based on MLPA and FISH assays.

Statistical softwares

Statistical analyses were performed using GraphPad Prism version 6, except for the 95% confidence intervals calculation and McNemar test, which were done using MedCalc Version 19.3.1 and IBM SPSS version 25, respectively.

Results

Assessment of the performance of designed MLPA probes in presence of normal genome

MLPA was selected as a suitable assay to measure CNAs from small amounts of genomic DNA that can be extracted from FFPE tumor samples in a multiplex format (Figure 1A). A custom-made MLPA probe mix including nine control probes targeting CNA quiet loci was designed to simultaneously determine the CNA status of ten genes relevant to PCa prognostication (Table 2). To assess potential probe-probe interactions, a MLPA reaction was carried out in the absence of DNA (water) and showed no peaks in the region corresponding to the designed probes (101 to 232 bp), confirming the absence of false-positive signals (Figure 1B). As expected, strong quantity control Q-DNA probe peaks at 64, 70, 76, and 82 bp, indicated the absence of DNA contamination and probe-probe interactions in an otherwise successful PCR reaction.

In presence of normal genomic DNA, each probe generated one specific fluorescence peak, all within the acceptable range of 1000 to 35000 relative fluorescent units (RFU), that appeared on capillary electrophoresis 2-4 nucleotides smaller in average than the designed length, which is consistent with the results obtained by Stern et al [33] and likely explainable by the negative charge-to-mass ratio of the dye [54] (Figure 1C). Consistent with normal copy number values, the probe ratios in the fresh normal genomic DNA (4 replicates of both male and female) and FFPE (6 replicates of both normal kidney and normal breast lymph node) were respectively between 0.94 and 1.04 and between 0.9 and 1.07 with less than 5% standard deviation (Figure 1D, Supplemental Table S1). The false-positive rate of the new assay was assessed in six separate experiments of MLPA (96 repeats each) using the PCa test probes mix with the designed reference probes or the commercial P200 – B1 reference probe mix on fresh normal genomic DNA. Each experiment assessed potential sources of variation due to the thermal cycler, the experimenter, the PCR plates, the caps and the presence of a mineral oil overlay to minimize evaporation. On average across the six experiments, the false CNA call was very low, representing 3.65% (95% confidence interval: 2.77% - 4.52%) of the reactions or 0.89% of the

probes in each experiment (total of 2,784 probes; 95% confidence interval: 0.46% - 1.31%). The error rate was independent of probes or any of the assessed parameters mentioned above and appeared to be intrinsic to the assay.

Detection of CNAs in PCa cell lines by MLPA

The PCa probe mix was next applied to fresh genomic DNA extracted from PCa cell lines (PC-3, DU145, LNCaP, and LAPC4). The Pearson correlation between the probe ratios of duplicate reactions was 0.985 (95% confidence interval: 0.978 – 0.989, p <0.0001). Reliability of the assay was assessed by repeating MLPA performed on PC-3 genome in ten separate experiments which showed consistent CNA calls (Fleiss' Kappa= 0.79, 95% confidence interval: 0.72 – 0.85, p <0.0001).

In all probes, with the exception of *CHD1* intron 1, a positive Pearson coefficient of correlation was seen when comparing the log2 transformed probe ratios generated from MLPA analysis of the four cell lines and those obtained from the two published array-CGH datasets of Zhao et al [52] and the cancer cell line encyclopedia (CCLE) [53] (Table 3). Of note, MLPA *CHD1* intron 1 probe targets the cytoband 5q21.1 while the exon probe targets 5q15 as the array-CGH probe. Moreover, the correlation coefficients between the MLPA ratios obtained from probes targeting the same gene were all above 0.7, with the exception of *PDPK1* (0.56). The correlation coefficients that are statistically significant are indicated in Table 3.

Considering the standard cutoff for gain and deletion, MLPA detected previously reported CNAs such as *MYC* gain, *PTEN* homozygous deletion, and *NKX3-1* (exon 2c probe) hemizygous deletion in PC-3 cells (Figure 2A, upper panel) as well as deletions of *CDKN1B* and *CHD1* (intron 1 probe) in LAPC4 cells (Figure 2A, lower panel). Of note, the *NKX3-1* and *CHD1* deletions were respectively identified in PC-3 and LAPC4 cells by both probes targeting each of these genes only when considering the 95% confidence interval. Moreover, the previously

reported hemizygous deletion of *TP53* in PC-3 cells [55] was called by the confidence interval approach, but completely missed when considering the fixed cutoff. These observations indicated that the sensitivity of the MLPA assay is increased by using the 95% confidence interval instead of a fixed cutoff for calling the CNAs.

Intriguingly, ratios of probes targeting exon 9 of *PTEN* and *NKX3-1* suggested a single copy gain of these genes in LAPC4 (Figure 2A, lower panel). To confirm these observations, the commercially available *PTEN* probe mix targeting exons 1 to 9 of the gene and flanking genes at 10q23.31 (MRC-Holland) was applied on both cell lines. As expected and in agreement with the literature [56], a deletion from exons 3 to 9 of *PTEN* was detected in PC-3 (Figure 2B, upper panel) and one copy gain (exons 1-9) was observed in LAPC4 (Figure 2B, lower panel).

Further validation was performed by FISH analysis on PC-3 and LAPC4 metaphase spreads with XT PTEN/GRID1 probe (Metasystems) that targets the distal region of the *PTEN* gene from exon 4 (red), the chromosome 10 centromere (aqua) and *GRID1* (green) located at 10q23.1 between *PTEN* and the centromere. In PC-3 cells, an average of 3.6 green and 4.6 aqua signals were detected per nucleus with no red signal, suggesting a *PTEN* deletion with a gain of chromosome 10 centromere (Figure 2C). In LAPC4, the 4.3 average red and green signals suggested a gain at 10q region including both the *PTEN* and *GRID1* genes as well as a chromosome 10 centromere duplication with an average of 3.3 aqua signal per nucleus (Figure 2C).

Using *MYC* targeting probes, FISH in PC-3 showed an amplification with an average of 12.1 (red) signals per nucleus along with a 2.9 (green) centromeric signals for chromosome 8 (Figure 2C). In LAPC4, FISH suggested a *MYC* and a chromosome 8 centromere amplification with an average of 3.9 signals per nucleus for both test and control probes (Figure 2C). For both genes and cell lines, the ratios of CNA probes over control probes were similar in MLPA and FISH. As

illustrated with the LAPC4 cells, MLPA, like other ratio-based CNA assessment method, can detect genomic imbalances, but balance ploidies result in normal probe ratios (undetected).

Assessment of the CNA detection limit

In the context of prostate tumor tissues heterogeneity and to estimate the minimum content of cancer cell DNA to accurately call CNAs with MLPA, a series of samples containing increasing percentages of PC-3 DNA (0-100%) over normal human male genomic DNA was used as a template for MLPA. Fresh female genome was used as the reference sample. Using pre-defined cutoffs to call CNAs, amplification of *MYC,* deletion of *PTEN* and *NKX3-1*, and one copy gain of *CHD1* (exon 35 probe) could be accurately detected when the PC-3 genome was at least at 10%, 60%, 90%, and 80%, respectively (Figure 3A-D). However, if the CNA calls are made based on the 95% confidence interval, all deletions and gains can be detected when the percentage of PC-3 genome is as low as 30% while *MYC* amplification is still detectable at 10% (Figure 3A-D). Since tumor samples could have variable tumor content with different levels of CNAs, a highly sensitive approach is required to accurately detect these alterations. This is especially important in prognostic assays on low-grade tumors that often have lower tumor contents and low CNA levels. The 95% confidence interval approach would thus appear more suitable to call CNAs in clinical samples.

Detection of CNAs using the MLPA designed probe mix in FFPE prostate tumor samples

MLPA was performed in duplicate reactions on all samples of the test sample set. To account for the range of variability expected in FFPE clinical samples and to match their DNA quality, we opted to include two FFPE extracted DNA references (normal breast and kidney) in addition to fresh genomic DNA. All reactions passed the QC and were included in the analysis. In parallel,

FISH for the genes assessed by MLPA was performed on all corresponding samples represented in a TMA.

Since each gene is targeted by two probes in the PCa probe mix and duplicate reactions were performed for each sample, various combinations of one or two probes in one or both replicates were considered to make the final CNA call for a given sample. To explore different normalization schemes for determining the CNA calls per reaction, we first normalized all reactions against the average ratio of the three DNA reference samples. In the second normalization scheme, we normalized all reactions against each of the three DNA reference samples separately, yielding three CNA data points per replicate reaction. We further explored whether conventional cutoff points, or the 95% confidence interval were best suited to call the CNA in each normalization scheme. Similar to the results obtained with cell lines, a lack of sensitivity of cutoff points in detection of CNAs called by FISH was evident (Supplemental Table S2, approaches 1 to 8). Based on the 95% confidence interval, the sensitivity of assays remained low when the data were normalized against the average value of the three reference samples (Supplemental Table S2, approaches 9 to 12). However, the sensitivity, based on the 95% confidence interval, was improved when the dataset was normalized against each reference sample separately and the alteration was detected in at least two of the three separately reference-normalized datasets (Supplemental Table S2, approaches 13-16). Although the sample size was limited, results appeared optimal for most CNAs with approach 14, yielding a median sensitivity of 80%, specificity of 93% and accuracy of 89% (Supplemental Table S2). With this approach, the data is normalized against each of the three reference DNA samples separately (Supplemental Figure S1). CNA in a gene was thus defined when both duplicate reactions show the same CNA in at least one of the probes targeting the gene and in at least two of the three normalizations. Figure 4A illustrates final MLPA CNA calls (based on approach 14) and the corresponding FISH CNA calls for the test sample set. An example of MLPA probe ratio profile

is shown in Figure 4B for sample (7-C) in which the deletions of *NKX3-1*, *PTEN* and *RB1* were also called by FISH (Figure 4C).

As illustrated in Figure 4A, the three benign samples were CNA quiet. While one of the benign samples (6-B) did not show any CNA, sample 1-B showed *GABRAPL2* deletion which was not seen in the corresponding tumor (1-C). The third benign sample (10-B) showed *RB1* deletion which was also detected in the corresponding cancer sample 10-C. However, no CNAs were detected by FISH in the benign samples.

In cancer samples, heterogeneity was evident among surveyed genes, ranging from none in sample 2-C to six in sample 13-C as measured by MLPA (Figure 4A). The highest CNA frequency were for *RB1* (9/15, 60% in both MLPA and FISH) and *NKX3-1* (7/15, 47% in both MLPA and FISH), whereas the lowest CNA frequency was for *PDPK1* (no CNA was detected with either method) and *TP53* (1/15, 7% in both methods). The data generated by the nine reference probes were examined for the presence of CNAs which were detected only for ANKRD36B_Int44 (3/18, 17%) and TIMM10_E3 (1/18, 6%, Supplemental Table S3).

To validate this analysis approach, the same experimental design was used to perform MLPA on genomic DNA extracted from 20 RP-FFPE independent cases in the validation sample set. Each case was represented by two samples, A and B corresponding to the highest and the lowest Gleason grade pattern, respectively. All samples passed the QC and were included in the analysis.

Given our laboratory interest in *PTEN* and *PDPK1* as prognostic biomarkers in PCa patients [18, 41], FISH was performed for these two genes on full sections of the FFPE blocks of this sample set targeting the area where tumor tissue cores were harvested for DNA extraction. Digital droplet TaqMan PCR assays were also performed to survey *PTEN* and *PDPK1* CNAs in 17 samples with the same DNA extracts used for MLPA (Supplemental Table S4). Figure 5A shows the CNA profile of the final call for MLPA, FISH and TaqMan. An example of the MLPA

profile for sample 23-C along with FISH for *PTEN* and *PDPK1* is presented in Figures 5B and 5C. As MLPA and FISH, digital droplet TaqMan PCR results confirmed the deletion of *PTEN* and gain of *PDPK1* with a probe ratio of 0.48 and 1.24, respectively (Supplemental Table S4). In this validation sample set, deletions of *NKX3-1* (25/40, 63%) and *RB1* (21/40, 53%) were the most frequent CNAs detected by MLPA, while *PDPK1* gain (5/40, 13%) was the least common, similar to results observed in the test set.

CNAs detected by the reference probes in the validation set were found mostly in ANKRD36B_Int44 (5/40, 13%), CYP2B6_Int2 (4/40, 10%) ATP10A_E20 (4/40, 10%) and TIMM10_E3 (3/40, 8%, Supplemental Table S5). The maximum number of reference genes affected by CNA per sample was two and observed in only two samples. Compared to FISH, the sensitivity and specificity of MLPA to call a *PTEN* deletion were respectively 85% and 100%. MLPA was able to detect *PDPK1* gain with a sensitivity of 100% and a specificity of 92%. Compared to TaqMan, MLPA showed a sensitivity of 100% and a specificity of 83% in detecting *PTEN* deletion and a sensitivity of 67% and a specificity of 93% in detection of *PDPK1* gain. However, if we consider the MLPA CNA calls based on the *PDPK1* intron 10 probe only, which assesses the same region targeted by the digital droplet TaqMan PCR *PDPK1* probe, the sensitivity and specificity to detect the *PDPK1* gain reach 100% and 93% respectively.

The impact of sampling was investigated by comparing the CNA patterns of tumor samples A and B taken from the same FFPE prostate block of each case. No significant differences in CNA calls was observed between samples A and B (McNemar's test $p > 0.4$, Supplemental Table S6).

Discussion

A detailed molecular classification of PCa and its associated prognostic value rely on the detection of multiple genomic alterations. FISH, the routine CNA detection method for tissue

sections, is not suitable for high-throughput assessment of CNA signatures. Standard array-CGH method and whole genome next generation sequencing (NGS) allow a genome-wide survey of CNAs but the amount and the quality of DNA required cannot be obtained from FFPE prostate biopsy specimens. Variations of the array-CGH method such as OncoScan microarray system [57] and targeted NGS [58] allow the call of CNAs out of small amounts of DNA extracted from FFPE tissue samples. These methods can be very comprehensive and well suited to explore the genomic landscape of tumors but would often require further confirmation for a definitive diagnosis. MLPA is a well-established method to call CNA widely used in human genetic test centres and considered as one of the gold standards which are used to validate new NGS-based tests [59]. In laboratories already offering genome-wide testing, MLPA assay may serve to rapidly confirm NGS and array-CGH findings. Moreover, the MLPA assay is flexible and new probes can be added or replaced to suit specific needs. For biomarker signatures application, a genome-wide survey is often not necessary and MLPA may be considered as a suitable and more affordable alternative to NGS and arrays-CGH allowing the analysis of large number of samples simultaneously. The low cost of MLPA can be particularly advantageous in the context of research work and clinical laboratory with limited resources. Compared to NGS, MLPA requires simpler analyses that translate into a faster turnaround time. Performed in a single PCR tube, MLPA requires only a thermal cycler and a standard capillary electrophoresis system. MLPA technology was thus selected for our new assay amenable to prostate biopsy material and filling the gap between biomarker research discovery and clinical application.

A synthetic probe mix was designed specifically for CNA detection in PCa targeting ten commonly altered genes that have the potential to serve as biomarker. This custom-made PCa specific probe mix was capable of accurately detecting CNAs in various samples with low error-rate. Using a synthetic DNA service with a high coupling efficiency of 99.5% (Ultramer DNA oligonucleotides, IDT) reduced the cost and time associated with the development of the PCa

probe mix while allowing an increased length of the synthetic probes from 156 nucleotides reported by Stern et al [33] to 232 nucleotides. The performance of these probes was validated by the reproducibility of the results obtained and the low standard deviation of all probes in the probe mix when using both fresh and FFPE extracted DNA. Our findings are in agreement with a previous report showing that synthetic probes gave comparable results to M13 bacteriophage derived probes used in commercially available probe mixes [33]. Based on our experience when using a coupling efficiency of 99.5%, synthetic MLPA probes of a length of up to 250 nucleotides might be used to generate highly reproducible results and allow the inclusion of a larger number of probes in the assay.

We further improved and customized the probe mix by designing nine different reference probes less likely to be affected by CNA based on reported PCa datasets. The rationale was to account for possible genomic instability of cancer cells that could result in random CNAs and the incompatibility of PCa samples with commercially available reference probes mixes such as P200. Although infrequent CNAs were detected by our reference probes, they did not compromise the analyses as there was an ample number of CNA-free references for the normalization of each clinical sample. Furthermore, as we and others have observed [33], a higher number of reference probes allows for better normalization, higher reproducibility of results and smaller standard deviation across different samples [33, 60]. Designing customized reference probes with an average distance of 4 nucleotides apart also allowed us to include more test probes in our PCa-specific probe mix than P200 probes, which have an average of 6 nucleotides apart and do not allow inclusion of more than 17 probes per reaction. Furthermore, for more accurate CNA detection, two probes per gene were used in the design process to assay each of the ten chosen target genes. The rationale stems from previous studies showing that MLPA probes targeting the same regions could yield different results [60, 61]. Bendavid *et al.* [62] used two different MLPA probe mix kits targeting the same regions and showed that not all alterations can be captured by

both kits. Alterations that were only seen by one of the kits were further confirmed by qPCR. Altogether, this supports that more than one probe per target region may allow better CNA detection, especially in tumor samples. The use of duplicate reactions in the experimental design was advantageous to reduce false-positive results and yielded better reproducibility. We reasoned that results would be reliable if the CNA was detected in both replicates. Further analysis confirmed high sensitivity and specificity of MLPA duplicate reactions in detection of CNAs compared to FISH data. In addition, a high correlation between duplicate reactions was observed in all experiments. The results also demonstrated that this MLPA assay performed in various experimental conditions yielded a lower rate of false-positive results with an average: 0.89% of probes used in each experiment compared to 1.7% reported with commercially available probe mixes [62].

To investigate the accuracy of CNA detection by the designed probe mix, we compared the CNA results obtained from MLPA to published array-CGH CNA results of PCa cell lines. Although these studies have used a different technique with different set of probes with a lower resolution than MLPA, our results are consistent with previous reports [60] and show a positive correlation between MLPA and array-CGH data, except for the *CHD1* intron 1 probe. *CHD1* is a large gene (74,569 bp) that spans over two cytobands (5q15-5q21.1). A CNA that would not span the entire *CHD1* gene in some cell lines such as PC-3 could potentially explain a lack of positive correlation of the MLPA intron probe (5q21.1) while the exon probe (5q15) correlates with array-CGH (5q15). Accordingly, the CNA would extend over both cytobands in LAPC4 cell line as both probes correlate. Testing this hypothesis could be done by mapping the CNA area in those cell lines with multiple MLPA probes. Nevertheless, both probes were effective at detecting 18 out of 20 *CHD1* deletions identified by MLPA in our clinical samples and the remaining two deletions were recognized by either the intron or the exon probe.

While these results indicate that MLPA is capable of detecting copy number imbalances in homogeneous samples such as cell lines, the accurate detection of genomic alterations in heterogeneous samples such as prostate primary tumors is more challenging. It is not specific to MLPA as most genomic assays that use DNA extracted from heterogeneous tumors face the same limitations. This was reported by Schwab *et al.* [60] to explain discrepancies observed between MLPA and FISH. The main reason for such discrepancies is intrinsic to the prostate *per se*, and the heterogeneous tumor content used for genomic DNA extraction which contain tumor DNA mixed with varying proportions of DNA from surrounding benign glands and stroma. Furthermore, when there is a low percentage of tumor cells with CNA in samples, due to the same dilution effect, genomic assays are not capable of accurately detecting alterations. To overcome this limitation, we used the 95% confidence interval for each probe within the reference population and test samples to detect CNAs. This approach has the advantage of calculating the range of variations seen for each locus in reference samples and detecting alterations outside what is normally seen in healthy tissues, thereby providing a more accurate measurement of CNAs. Thus, using the confidence interval approach, we were able to capture CNAs when the PC-3 genome was diluted up to 30%. When compared to FISH, this method of analysis of MLPA data on the test clinical samples resulted in a median sensitivity of 80% and specificity of 93% compared to the conventional cutoff points approach with the sensitivity of 14% and specificity of 100%. Even when analyzing cell lines, we noticed that when using cutoff points, the previously reported CNAs such as deletion of *TP53* in PC-3 was completely missed or called only by one of the two probes for the deletion of *NKX3-1* in PC-3 and *CHD1* in LAPC4. However, when using the 95% confidence interval, all these CNAs were readily identified which further supports the use of this approach over fixed cutoff points.

In both test and validation sets, the number of CNAs varied across the samples as reported in genomic profiling studies [15, 26]. The most frequent CNAs detected in tumors by both MLPA and

FISH were *RB1* and *NKX3-1* deletion, a finding in agreement with previous literature [15] and consistent with reports suggesting that inactivation of these two tumor suppressors occurs early in prostate carcinogenesis [63, 64]. Two out of three benign prostate samples showed a single CNA each by MLPA, respectively a *GABARAPL2* and a *RB1* deletion, which were not detected by FISH. The benign samples of this study were all from PC patients and the *RB1* deletion was also detected in the corresponding tumor sample. Although false positive results cannot be excluded, genomic alterations have been detected in morphologically benign prostate tissues that were compatible with field effects in a context of multifocal PCa [65]. Profiling additional samples taken from multiple areas of tumoral and corresponding benign prostate tissues as well as benign samples from non-cancer patients would be warranted to confirm this hypothesis.

When comparing the results of MLPA and FISH in the test and validation samples, we observed good concordance between the two types of assays. Yet discrepancies exist in some prostate samples, with MLPA revealing CNAs but not FISH and vice versa. The dilution effect can reduce the sensitivity of MLPA as discussed above and unsuspected DNA sequence variations at the probe hybridizing site may lead to MLPA false positive results [66]. However, the size and the location of the targeted DNA region by FISH and MLPA probes may account for the discrepancies between both techniques. MLPA with its small probe size allows the mapping of CNAs at higher resolution than large FISH probes. Including more MLPA probes per gene or using FISH probes that target different areas of the CNA region can help resolve the observed discrepancies.

To circumvent any differences linked to probe genomic location and to tumor areas, digital droplet PCR TaqMan assays targeting the same genomic regions and performed on the same DNA extracts showed a high concordance with MLPA results. This further showed the robustness of the designed assay. It is worth emphasizing that TaqMan has limited multiplexing

capability and, in this case, only one reference probe was used, whereas the designed MLPA assay used nine different reference probes, thereby allowing more accurate normalization.

We also assessed CNA frequencies in different Gleason grade patterns of the same samples in our small validation sample set and did not observe significant differences, thereby suggesting negligible intra-tumor CNA heterogeneity in the assessed genes in primary tumors. A more robust study with a larger sample size that would include a wider range of Gleason patterns is needed to confirm these results.

Taken together, we have designed and rigorously tested a MLPA-based assay using a fully synthetic probe mix. DNA extracts from three 0.6 mm punches (2-3 mm length) from FFPE clinical samples were used to simulate needle biopsies, as obtained before diagnosis. The assay was capable of detecting CNAs with high sensitivity and specificity compared to FISH. This demonstrates the potential compatibility and applicability of the assay on clinical biopsy samples for the detection of most common CNAs, thereby opening the door to a better prognosis based on CNA profiles. Future investigations should aim to confirm these findings on actual prostate diagnostic biopsies and may require a macrodissection of the tumor foci from nearby normal glands since the tumor cell content is expected to be less than in RP blocks used for this study. Furthermore, the cost is low, our estimate being 7.90 $CAD (6$ USD) per reaction which includes the cost of probe synthesis, lab materials, MLPA kits, and capillary electrophoresis. MLPA does not require other sophisticated instruments or analysis approaches. The effectiveness of this designed probe mix and analysis approach was validated using a small number of clinical PCa samples. Further validation on a larger cohort with adequate clinical follow-up data is necessary to determine the impact of these CNAs and utility of the assay for prognosis. This assay shows potential for clinical application, and by identifying CNAs relevant to poor prognosis could aid clinicians and patients in decision-making in view of an optimal management of the disease.

Acknowledgments

We would like to thank patients who consented to provide prostate tissues for research and Dr Jose Correa from the Department of Mathematics and Statistics, McGill University for statistical consultation. We are also grateful to Dr Amjad Alwaal and Mr Joshua Ejdelman for their early work on the selection and validation of some of the FISH probes used for this research. We also thank Dr Josée Lavoie from the Molecular Genetics Laboratory, McGill University Health Centre for providing normal metaphase chromosome preparations and assistance with the implementation of the FISH method.

Reference

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2019. CA: A Cancer Journal for Clinicians 2019, 69:7-34.
2. Canadian Cancer Statistics Advisory Committee: Canadian Cancer Statistics 2018. Toronto, ON: Canadian Cancer Society, 2018.
3. Klotz L, Vesprini D, Sethukavalan P, Jethava V, Zhang L, Jain S, Yamamoto T, Mamedov A, Loblaw A: Long-Term Follow-Up of a Large Active Surveillance Cohort of Patients With Prostate Cancer. Journal of Clinical Oncology 2014, 33:272-277.
4. Litwin MS, Tan H-J: The Diagnosis and Treatment of Prostate Cancer: A ReviewThe Diagnosis and Treatment of Prostate CancerThe Diagnosis and Treatment of Prostate Cancer. JAMA 2017, 317:2532-2542.
5. Hutchinson L: Closing the controversies gap in prostate cancer? Nature reviews 2014, 11:299.
6. Lotan TL, Heumann A, Rico SD, Hicks J, Lecksell K, Koop C, Sauter G, Schlomm T, Simon R: PTEN loss detection in prostate cancer: comparison of PTEN immunohistochemistry and PTEN FISH in a large retrospective prostatectomy cohort. Oncotarget 2017, 8:65566-65576.
7. Shore ND, Kella N, Moran B, Boczko J, Bianco FJ, Crawford ED, Davis T, Roundy KM, Rushton K, Grier C, Kaldate R, Brawer MK, Gonzalgo ML: Impact of the Cell Cycle Progression Test on Physician and Patient Treatment Selection for Localized Prostate Cancer. Journal of Urology 2016, 195:612-618.
8. Knezevic D, Goddard AD, Natraj N, Cherbavaz DB, Clark-Langone KM, Snable J, Watson D, Falzarano SM, Magi-Galluzzi C, Klein EA, Quale C: Analytical validation of

the Oncotype DX prostate cancer assay – a clinical RT-PCR assay optimized for prostate needle biopsies. BMC Genomics 2013, 14:690.

9. Spratt DE, Yousefi K, Deheshi S, Ross AE, Den RB, Schaeffer EM, Trock BJ, Zhang J, Glass AG, Dicker AP, Abdollah F, Zhao SG, Lam LLC, Plessis Md, Choeurng V, Haddad Z, Buerki C, Davicioni E, Weinmann S, Freedland SJ, Klein EA, Karnes RJ, Feng FY: Individual Patient-Level Meta-Analysis of the Performance of the Decipher Genomic Classifier in High-Risk Men After Prostatectomy to Predict Development of Metastatic Disease. Journal of Clinical Oncology 2017, 35:1991-1998.
10. Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. Nature Reviews Genetics 2006, 7:85.
11. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M: The landscape of somatic copy-number alteration across human cancers. Nature 2010, 463:899.
12. Kim T-M, Xi R, Luquette LJ, Park RW, Johnson MD, Park PJ: Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. Genome research 2013, 23:217-227.
13. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, Shiah Y-J, Yousif F, Lin X, Masella AP: Genomic hallmarks of localized, non-indolent prostate cancer. Nature 2017, 541:359.
14. Espiritu SMG, Liu LY, Rubanova Y, Bhandari V, Holgersen EM, Szyca LM, Fox NS, Chua ML, Yamaguchi TN, Heisler LE: The evolutionary landscape of localized prostate cancers drives clinical aggression. Cell 2018, 173:1003-1013. e1015.
15. Lapointe J, Li C, Giacomini CP, Salari K, Huang S, Wang P, Ferrari M, Hernandez-Boussard T, Brooks JD, Pollack JR: Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. Cancer research 2007, 67:8504-8510.

16. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, Landers T, Dolgalev I, Major JE, Wilson M, Socci ND, Lash AE, Heguy A, Eastham JA, Scher HI, Reuter VE, Scardino PT, Sander C, Sawyers CL, Gerald WL: Integrative genomic profiling of human prostate cancer. Cancer Cell 2010, 18:11-22.

17. Barros-Silva JD, Ribeiro FR, Rodrigues A, Cruz R, Martins AT, Jeronimo C, Henrique R, Teixeira MR: Relative 8q gain predicts disease-specific survival irrespective of the TMPRSS2-ERG fusion status in diagnostic biopsies of prostate cancer. Genes Chromosomes Cancer 2011, 50:662-671.

18. Bramhecha YM, Rouzbeh S, Guérard K-P, Scarlata E, Brimo F, Chevalier S, Hamel L, Aprikian AG, Lapointe J: The combination of PTEN deletion and 16p13. 3 gain in prostate cancer provides additional prognostic information in patients treated with radical prostatectomy. Modern Pathology 2019, 32:128.

19. Kluth M, Harasimowicz S, Burkhardt L, Grupp K, Krohn A, Prien K, Gjoni J, Hass T, Galal R, Graefen M, Haese A, Simon R, Huhne-Simon J, Koop C, Korbel J, Weischenfeld J, Huland H, Sauter G, Quaas A, Wilczak W, Tsourlakis MC, Minner S, Schlomm T: Clinical significance of different types of p53 gene alteration in surgically treated prostate cancer. International journal of cancer 2014, 135:1369-1380.

20. Krohn A, Diedler T, Burkhardt L, Mayer PS, De Silva C, Meyer-Kornblum M, Kotschau D, Tennstedt P, Huang J, Gerhauser C, Mader M, Kurtz S, Sirma H, Saad F, Steuber T, Graefen M, Plass C, Sauter G, Simon R, Minner S, Schlomm T: Genomic deletion of PTEN is associated with tumor progression and early PSA recurrence in ERG fusion-positive and fusion-negative prostate cancer. The American journal of pathology 2012, 181:401-412.

21. Qian J, Hirasawa K, Bostwick DG, Bergstralh EJ, Slezak JM, Anderl KL, Borell TJ, Lieber MM, Jenkins RB: Loss of p53 and c-myc overrepresentation in stage T(2-3)N(1-3)M(0) prostate cancer are potential markers for cancer progression. Mod Pathol 2002, 15:35-44.

22. Yoshimoto M, Cunha IW, Coudry RA, Fonseca FP, Torres CH, Soares FA, Squire JA: FISH analysis of 107 prostate cancers shows that PTEN genomic deletion is associated with poor clinical outcome. British journal of cancer 2007, 97:678-685.

23. Zafarana G, Ishkanian AS, Malloff CA, Locke JA, Sykes J, Thoms J, Lam WL, Squire JA, Yoshimoto M, Ramnarine VR, Meng A, Ahmed O, Jurisca I, Milosevic M, Pintilie M, van der Kwast T, Bristow RG: Copy number alterations of c-MYC and PTEN are prognostic factors for relapse after prostate cancer radiotherapy. Cancer 2012, 118:4053-4062.

24. Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, Heguy A, Huberman K, Bernstein M, Assel M, Murali R, Vickers A, Scardino PT, Sander C, Reuter V, Taylor BS, Sawyers CL: Copy number alteration burden predicts prostate cancer relapse. Proc Natl Acad Sci U S A 2014, 111:11139-11144.

25. Kluth M, Runte F, Barow P, Omari J, Abdelaziz ZM, Paustian L, Steurer S, Christina Tsourlakis M, Fisch M, Graefen M, Tennstedt P, Huland H, Michl U, Minner S, Sauter G, Simon R, Adam M, Schlomm T: Concurrent deletion of 16q23 and PTEN is an independent prognostic feature in prostate cancer. International journal of cancer 2015, 137:2354-2363.

26. Lalonde E, Ishkanian AS, Sykes J, Fraser M, Ross-Adams H, Erho N, Dunning MJ, Halim S, Lamb AD, Moon NC, Zafarana G, Warren AY, Meng X, Thoms J, Grzadkowski MR, Berlin A, Have CL, Ramnarine VR, Yao CQ, Malloff CA, Lam LL, Xie H, Harding NJ, Mak DY, Chu KC, Chong LC, Sendorek DH, P'ng C, Collins CC,

Squire JA, Jurisica I, Cooper C, Eeles R, Pintilie M, Dal Pra A, Davicioni E, Lam WL, Milosevic M, Neal DE, van der Kwast T, Boutros PC, Bristow RG: Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. The Lancet 2014, 15:1521-1532.

27. Ceulemans S, van der Ven K, Del-Favero J: Targeted screening and validation of copy number variations. Genomic Structural Variants: Springer, 2012. pp. 311-328.

28. Ooi A, Inokuchi M, Horike S-i, Kawashima H, Ishikawa S, Ikeda H, Nakamura R, Oyama T, Dobash Y: Amplicons in breast cancers analyzed by multiplex ligation-dependent probe amplification and fluorescence in situ hybridization. Human pathology 2019, 85:33-43.

29. Dalay N: Detection of RB1 Gene Copy Number Variations Using a Multiplex Ligation-Dependent Probe Amplification Method. The Retinoblastoma Protein. Edited by Santiago-Cardona PG. New York, NY: Springer New York, 2018. pp. 7-18.

30. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE: The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours. Eur Urol 2016, 70:106-119.

31. Patel PG, Selvarajah S, Guérard K-P, Bartlett JM, Lapointe J, Berman DM, Okello JB, Park PC: Reliability and performance of commercial RNA and DNA extraction kits for FFPE tissue cores. PloS one 2017, 12:e0179732.

32. D'Angelo CS, Varela MC, de Castro CI, Kim CA, Bertola DR, Lourenço CM, Perez ABA, Koiffmann CP: Investigation of selected genomic deletions and duplications in a cohort of 338 patients presenting with syndromic obesity by multiplex ligation-dependent probe amplification using synthetic probes. Mol Cytogenet 2014, 7:75-75.

33. Stern RF, Roberts RG, Mann K, Yau SC, Berg J, Ogilvie CM: Multiplex ligation-dependent probe amplification using a completely synthetic probe set. Biotechniques 2004, 37:399-405.

34. Barbaro M, Cicognani A, Balsamo A, Löfgren Å, Baldazzi L, Wedell A, Oscarson M: Gene dosage imbalances in patients with 46,XY gonadal DSD detected by an in-house-designed synthetic probe set for multiplex ligation-dependent probe amplification analysis. Clinical Genetics 2008, 73:453-464.

35. Choucair K, Ejdelman J, Brimo F, Aprikian A, Chevalier S, Lapointe J: PTEN genomic deletion predicts prostate cancer recurrence and is associated with low AR expression and transcriptional activity. BMC cancer 2012, 12:543.

36. Hamid AA, Gray KP, Shaw G, MacConaill LE, Evan C, Bernard B, Loda M, Corcoran NM, Van Allen EM, Choudhury AD, Sweeney CJ: Compound Genomic Alterations of TP53, PTEN, and RB1 Tumor Suppressors in Localized and Metastatic Prostate Cancer. European Urology 2018.

37. Kluth M, Ahrary R, Hube-Magg C, Ahmed M, Volta H, Schwemin C, Steurer S, Wittmer C, Wilczak W, Burandt E: Genomic deletion of chromosome 12p is an independent prognostic marker in prostate cancer. Oncotarget 2015, 6:27966.

38. Kibel AS, Schutte M, Kern SE, Isaacs WB, Bova GS: Identification of 12p as a region of frequent deletion in advanced prostate cancer. Cancer research 1998, 58:5652-5655.

39. Kluth M, Scherzai S, Buschek F, Fraune C, Moller K, Hoflmayer D, Minner S, Gobel C, Moller-Koop C, Hinsch A, Neubauer E, Tsourlakis MC, Sauter G, Heinzer H, Graefen M, Wilczak W, Luebke AM, Burandt E, Steurer S, Schlomm T, Simon R: 13q deletion is linked to an adverse phenotype and poor prognosis in prostate cancer. Genes Chromosomes Cancer 2018, 57:504-512.

40. Choucair KA, Guerard KP, Ejdelman J, Chevalier S, Yoshimoto M, Scarlata E, Fazli L, Sircar K, Squire JA, Brimo F, Cunha IW, Aprikian A, Gleave M, Lapointe J: The 16p13.3 (PDPK1) Genomic Gain in Prostate Cancer: A Potential Role in Disease Progression. Translational oncology 2012, 5:453-460.

41. Bramhecha YM, Guérard K-P, Rouzbeh S, Scarlata E, Brimo F, Chevalier S, Hamel L, Dragomir A, Aprikian AG, Lapointe J: Genomic Gain of 16p13. 3 in Prostate Cancer Predicts Poor Clinical Outcome after Surgical Intervention. Molecular Cancer Research 2018, 16:115-123.

42. Burkhardt L, Fuchs S, Krohn A, Masser S, Mader M, Kluth M, Bachmann F, Huland H, Steuber T, Graefen M, Schlomm T, Minner S, Sauter G, Sirma H, Simon R: CHD1 is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer. Cancer research 2013, 73:2795-2805.

43. Huang S, Gulzar ZG, Salari K, Lapointe J, Brooks JD, Pollack JR: Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness. Oncogene 2012.

44. Rodrigues LU, Rider L, Nieto C, Romero L, Karimpour-Fard A, Loda M, Lucia MS, Wu M, Shi L, Cimic A, Sirintrapun SJ, Nolley R, Pac C, Chen H, Peehl DM, Xu J, Liu W, Costello JC, Cramer SD: Coordinate loss of MAP3K7 and CHD1 promotes aggressive prostate cancer. Cancer research 2015, 75:1021-1034.

45. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, Varambally S, Cao X, Tchinda J, Kuefer R: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. science 2005, 310:644-648.

46. Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera J-M, Setlur S, Tchinda J, Tomlins SA, Hofer MD, Pienta KG: TMPRSS2: ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer research 2006, 66:8337-8341.

47. Demichelis F, Fall K, Perner S, Andrén O, Schmidt F, Setlur S, Hoshida Y, Mosquera J, Pawitan Y, Lee C: TMPRSS2: ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene 2007, 26:4596.

48. Zhi J, Hatchwell E: Human MLPA Probe Design (H-MAPD): a probe design tool for both electrophoresis-based and bead-coupled human multiplex ligation-dependent probe amplification assays. BMC genomics 2008, 9:407.

49. Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, Parton A, Armean IM, Trevanion SJ, Flicek P, Cunningham F: Ensembl variation resources. Database 2018, 2018.

50. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013, 6:pl1-pl1.

51. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. AACR, 2012.

52. Zhao H, Kim Y, Wang P, Lapointe J, Tibshirani R, Pollack JR, Brooks JD: Genome-wide characterization of gene expression variations and DNA copy number changes in prostate cancer cell lines. The Prostate 2005, 63:187-197.

53. Research BIatNIfB: CCLE (Cancer Cell Line Encyclopedia). 2018.

54. Tu O, Knott T, Marsh M, Bechtol K, Harris D, Barker D, Bashkin J: The influence of fluorescent dye structure on the electrophoretic mobility of end-labeled DNA. Nucleic Acids Research 1998, 26:2797-2802.

55. Carroll AG, Voeller HJ, Sugars L, Gelmann EP: p53 oncogene mutations in three human prostate cancer cell lines. The Prostate 1993, 23:123-134.

56. Whang YE, Wu X, Suzuki H, Reiter RE, Tran C, Vessella RL, Said JW, Isaacs WB, Sawyers CL: Inactivation of the tumor suppressor PTEN/MMAC1 in advanced human prostate cancer through loss of expression. Proceedings of the National Academy of Sciences 1998, 95:5246-5250.

57. Jung H-S, Lefferts JA, Tsongalis GJ: Utilization of the oncoscan microarray assay in cancer diagnostics. Applied Cancer Research 2017, 37:1.

58. Seed G, Yuan W, Mateo J, Carreira S, Bertan C, Lambros M, Boysen G, Ferraldeschi R, Miranda S, Figueiredo I, Riisnaes R, Crespo M, Rodrigues DN, Talevich E, Robinson DR, Kunju LP, Wu YM, Lonigro R, Sandhu S, Chinnaiyan AM, de Bono JS: Gene Copy Number Estimation from Targeted Next-Generation Sequencing of Prostate Cancer Biopsies: Analytic Validation and Clinical Qualification. Clin Cancer Res 2017, 23:6070-6077.

59. Kerkhof J, Schenkel LC, Reilly J, McRobbie S, Aref-Eshghi E, Stuart A, Rupar CA, Adams P, Hegele RA, Lin H, Rodenhiser D, Knoll J, Ainsworth PJ, Sadikovic B: Clinical Validation of Copy Number Variant Detection from Targeted Next-Generation Sequencing Panels. J Mol Diagn 2017, 19:905-920.

60. Schwab C, Jones L, Morrison H, Ryan S, Yigittop H, Schouten J, Harrison C: Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia. Genes, Chromosomes and Cancer 2010, 49:1104-1113.

61. Marcinkowska-Swojak M, Uszczynska B, Figlerowicz M, Kozlowski P: An MLPA-Based Strategy for Discrete CNV Genotyping: CNV-mi RNA s as an Example. Human mutation 2013, 34:763-773.

62. Bendavid C, Dubourg C, Pasquier L, Gicquel I, Le Gallou S, Mottier S, Durou MR, Henry C, Odent S, David V: MLPA screening reveals novel subtelomeric rearrangements in holoprosencephaly. Human mutation 2007, 28:1189-1197.

63. Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, Kim M, Desai N, Young P, Norton CR, Gridley T, Cardiff RD, Cunha GR: Roles for Nkx3. 1 in prostate development and cancer. Genes & development 1999, 13:966-977.

64. Maddison LA, Sutherland BW, Barrios RJ, Greenberg NM: Conditional deletion of Rb causes early stage prostate cancer. Cancer research 2004, 64:6018-6025.

65. Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, Kremeyer B, Butler A, Lynch AG, Camacho N: Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. Nature genetics 2015, 47:367.

66. Vijzelaar R, Botton MR, Stolk L, Martis S, Desnick RJ, Scott SA: Multi-ethnic SULT1A1 copy number profiling with multiplex ligation-dependent probe amplification. Pharmacogenomics 2018, 19:761-770.

Figure Legends

Figure 1: Performance of the newly designed MLPA probes in presence of normal genomic DNA. (**A**) Schematic representation of MLPA reaction. No annealing takes place in absence of target DNA (left panel) while in presence of the target sequence, the two MLPA half-probes anneal to the DNA next to each other (middle and right panels). After a ligation step, the complete probes are amplified using fluorescently labeled primers. The amount of fluorescent signal is correlated with the copy number of the target sequence. (**B, C**) Electropherogram of PCa probe mix in (**B**) no-DNA control reaction and in (**C**) presence of normal genome. Relative fluorescent units (RFU) are shown for probes from left to right according to their length. Probes are identified at the top of their respective peak. (**D**) MLPA profile of normalized probe ratios in normal genome. The upper blue line is the cut-off for gain and the red for deletions set at 1.3 and 0.65, respectively. Reference probes are marked by an **asterisk** at the top of the probe name.

Figure 2: Detection of CNAs in PC-3 and LAPC4 cell lines. (**A**, **B**) In MLPA ratio charts, probes in blue and red correspond to gain and deletion based on the standard cut-off point of above 1.3 and below 0.65, respectively. Probes in purple show gain or deletion based on the 95% confidence interval only. Reference probes are marked by an **asterisk** at the top of the probe name. (**A**) MLPA results of PC-3 (upper panel) and LAPC4 (lower panel) cell lines using the designed PCa probe mix. The red and the blue box highlights respectively the *PTEN* deletion in PC-3 and the *PTEN* gain in LAPC4. (**B**) As in (**A)**, but showing the MLPA profiles of PC-3 and LAPC4 cell lines using commercial *PTEN* probe mix (P225-PTEN, D2-0315, MRC-Holland). (**C**) FISH results of *PTEN* and *MYC* on metaphase spreads of PC-3 and LAPC4 cell lines. The *PTEN* FISH probe targets three regions of chromosome 10 (XT PTEN/GRID1 probe, Metasystems). The red signal corresponds to the *PTEN* gene, the green signal represents the *GRID1* gene while the aqua signal is the centromere control. The *MYC* fluorescent signal is red

and the centromere control for chromosome 8 is green. The average number of signals detected and the test probe over control probe counts ratio is indicated on each FISH picture (96x magnification).

Figure 3: Using 95% confidence interval of probes improves CNA detection limit of MLPA. (**A**, **B**, **C**, **D**) Probe ratio values are plotted against percentage of PC-3 genome for (**A**) *MYC*, (**B**) *PTEN*, (**C**) *NKX3-1* and (**D**) *CHD1*. Dashed lines show the standard cut-off points for gains and deletions based on probe ratios at 1.3 and 0.65, respectively. The solid lines show the cut-off points for gains and deletions based on 95% confidence interval of the probes in the reference DNA.

Figure 4: CNA profiles of the test sample set. (**A**) Final MLPA and FISH CNA calls in the test sample set. Patient ID is followed by C for cancer or B for benign sample. (**B**) Example of MLPA results for one sample (7-C) which shows deletions of *NKX3-1*, *PTEN* and *RB1*. Probes in red show deletion, according to the 95% confidence interval method. Reference probes are marked by an **asterisk** at the top of the probe name. (**C**) FISH results for sample 7-C showing gene specific signal in red and control signal in green for the 10 loci assessed. Deletions in *NKX3-1* and *RB1* (hemizygous) and *PTEN* (homozygous) are indicated with arrows (96x magnification).

Figure 5: CNA calls in the validation sample set. (**A**) Final MLPA CNA call for all genes assessed along FISH and digital droplet TaqMan PCR (a subset of samples) results for *PTEN* and *PDPK1* genes. Patient ID is followed by C (cancer) and either A or B (highest or lowest Gleason pattern area of the tumor, respectively). (**B**) MLPA profile of sample 23-C-A of the validation sample set showing deletions in *NKX3-1*, *PTEN*, *GABARAPL2* and *TP53* in red, and gain of

MYC and *PDPK1* in blue, according to the 95% confidence interval method. Reference probes are marked by an **asterisk** at the top of the probe name. (**C**) Representative images of FISH results for sample 23-C-A confirming the *PTEN* deletion and *PDPK1* gain detected by MLPA. **White arrows** show control green signals in absence and presence of multiple gene specific red signal in nuclei with *PTEN* deletion and *PDPK1* gain, respectively (96x magnification).

Table 1: List and characteristics of samples used in this study

ID	Type of the Core	Reviewed Gleason Grade of the Case	% of Cancer Cells	Sample Set
1-C	Cancer	4+5	85	Test
1-B	Benign	N/A	0	Test
2-C	Cancer	5+4	85	Test
3-C	Cancer	4+3	50	Test
4-C	Cancer	3+4	85	Test
5-C	Cancer	4+5	90	Test
6-C	Cancer	4+3	70	Test
6-B	Benign	N/A	0	Test
7-C	Cancer	4+5	70	Test
8-C	Cancer	3+4(5)	75	Test
9-C	Cancer	3+4	85	Test
10-C	Cancer	4+3(5)	90	Test
10-B	Benign	N/A	0	Test
12-C	Cancer	4+3(5)	90	Test
13-C	Cancer	4+3(5)	85	Test
14-C	Cancer	4+5	80	Test
16-C	Cancer	5+4	80	Test
18-C	Cancer	4+5	80	Test
19-C	Cancer	4+3(5)	95	Validation
20-C	Cancer	4+4	90	Validation
21-C	Cancer	4+5	90	Validation
22-C	Cancer	4+5	90	Validation
23-C	Cancer	4+4	75	Validation
24-C	Cancer	4+3	75	Validation
25-C	Cancer	4+4	90	Validation
26-C	Cancer	4+5(3)	85	Validation
27-C	Cancer	4+3	90	Validation
28-C	Cancer	3+4	70	Validation
29-C	Cancer	4+5	85	Validation
30-C	Cancer	4+3	80	Validation
31-C	Cancer	3+4	80	Validation
32-C	Cancer	3+4	65	Validation
33-C	Cancer	4+5	90	Validation
34-C	Cancer	3+4	80	Validation
35-C	Cancer	4+3(5)	70	Validation
36-C	Cancer	4+3	80	Validation
37-C	Cancer	4+3(5)	90	Validation
38-C	Cancer	4+3	80	Validation

Table 2: List and characteristics of designed PCa specific MLPA probe mix

Name	Function	Cytoband	Hybridization Sequence Coordinates (hg38)	Probe Length
CHD1 Intron 1	CNA	05q15	98927383-98927458	200
CHD1 Exon 35	CNA	05q21.1	98856501-98856587	132
MAP3K7 Exon 17	CNA	06q15	90516154-90516215	125
MAP3K7 Exon 14	CNA	06q15	90523700-90523761	164
NKX3-1 Exon 2c	CNA	08p21.2	23679195-23679260	188
NKX3-1 Exon 2	CNA	08p21.2	23679829-23679894	152
MYC Exon 1	CNA	08q24.21	127736564-127736621	101
MYC Exon 3	CNA	08q24.21	127740610-127740671	184
PTEN Exon 9a	CNA	10q23.31	87966936-87966999	109
PTEN Exon 9b	CNA	10q23.31	87968061-87968130	140
CDKN1B Intron 1	CNA	12p13.1	12714777-12714852	156
CDKN1B Exon 1	CNA	12p13.1	12717331-12717394	180
RB1 Exon 18	CNA	13q14.2	48453031-48453090	105
RB1 Exon 23	CNA	13q14.2	48465212-48465273	172
PDPK1 Exon 14	CNA	16p13.3	2597685-2597741	121
PDPK1 Intron 10	CNA	16p13.3	2584302-2584368	220
GABARAPL2 Exon 3	CNA	16q23.1	75568131-75568188	168
GABARAPL2 Intron 3	CNA	16q23.1	75573764-75573839	196
TP53 Exon 5	CNA	17p13.1	7674901-7674962	136
TP53 Exon 4	CNA	17p13.1	7675285-7675346	117
ANKRD36B Intron 44	Reference	02q11.2	97495557-97495632	232
MGAT1 Exon 3	Reference	05q35.3	180790824-180790885	160
TIMM10 Exon 3	Reference	11q12.1	57528778-57528839	204
METTL1 Intron 1	Reference	12q14.1	57774661-57774736	144
IPO4 Exon 30	Reference	14q12	24180336-24180397	176
ATP10A Exon 20	Reference	15q12	25680143-25680204	192
PIGW Intron 1	Reference	17q12	36533015-36533090	129
ZNF91 Intron 4	Reference	19p12	23350707-23350774	113
CYP2B6 Intron 2	Reference	19q13.2	41004294-41004359	216

Table 3: Correlation of log_2 transformed probe ratios between PCa specific MLPA probe mix, and array-CGH studies

Gene	Method	Study/Probe	Log_2 Transformed Normalized Probe Ratio				r with Zhao et al [52]	r with with CCLE [53]	r between MLPA Probes †
			PC-3	DU145	LNCaP	LAPC4			
CHD1	Array-CGH	Zhao et al [52] (exon 35)	0.81	-0.42	-1.07	-0.48			
		CCLE [53]	0.49	0.17	0.09	N/A	0.99		
	MLPA	*CHD1* Intron 1	-0.12	0.03	0.27	-0.96	-0.13	-0.89	
		CHD1 Exon 35	0.44	0.07	0.16	-0.43	0.50	0.91	0.78
MAP3K7	Array-CGH	Zhao et al [52] (exon 4)	-0.67	-0.35	-1.40	-0.45			
		CCLE [53]	-0.50	-0.29	-1.21	N/A	1.00*		
	MLPA	*MAP3K7* Exon 14	-0.18	-0.54	-0.55	-0.20	0.40	0.32	
		MAP3K7 Exon 17	-0.07	-0.55	-0.48	-0.02	0.30	0.16	0.98*
NKX3-1	Array-CGH	Zhao et al [52] (exon 2)	-0.40	-0.30	0.17	-0.06			
		CCLE [53]	-0.47	0.12	0.09	N/A	0.60		
	MLPA	*NKX3-1* Exon 2	-0.56	0.04	0.19	0.43	0.72	0.97	
		NKX3-1 Exon 2c	-0.73	0.19	0.28	0.27	0.71	0.99	0.94
MYC	Array-CGH	Zhao et al [52] (exon 3)	0.69	-0.10	-0.01	-0.17			
		CCLE [53]	0.95	0.32	0.09	N/A	0.93		
	MLPA	*MYC* Exon 1	0.84	0.23	0.17	0.29	0.95*	0.98	
		MYC Exon 3	3.14	0.19	-0.14	-0.20	0.98*	0.99	0.98*
PTEN	Array-CGH	Zhao et al [52] (exon 1)	-0.36	-0.03	-0.55	0.05			
		CCLE [53]	-2.40	0.15	-0.91	N/A	0.55		
	MLPA	*PTEN* Exon 9a	-3.74	-0.12	-2.65	0.59	0.88	0.95	
		PTEN Exon 9b	-5.64	0.01	-0.50	0.47	0.45	0.94	0.82
CDKN1B	Array-CGH	Zhao et al [52] (exon 3)	-0.29	-0.54	-0.13	-0.73			
		CCLE [53]	-0.51	-0.29	0.06	N/A	0.50		
	MLPA	*CDKN1B* Intron 1	-0.47	-0.34	0.19	-0.75	0.84	0.98	
		CDKN1B Exon 1	-0.55	-0.53	0.09	-0.92	0.90	0.93	0.99*
RB1	Array-CGH	Zhao et al [52] (exon 27)	0.23	0.15	0.15	0.29			
		CCLE [53]	-0.05	-0.31	-0.23	N/A	0.95		
	MLPA	*RB1* Exon 18	0.03	-0.22	-0.14	0.30	0.98*	1.00*	
		RB1 Exon 23	-0.14	-0.67	-0.35	-0.18	0.76	0.94	0.77
PDPK1	Array-CGH	Zhao et al [52] (exon 14)	0.34	0.12	0.27	0.16			
		CCLE	-0.53	-0.35	0.07	N/A	-0.09		
	MLPA	*PDPK1* Intron 10	-0.11	-0.22	0.17	-0.24	0.57	0.84	
		PDPK1 Exon 14	0.04	-0.32	0.15	0.08	0.62	0.44	0.57
GABARAPL2	Array-CGH	Zhao et al [52] (exon 4)	-0.10	0.02	-0.07	0.05			
		CCLE [53]	-0.50	0.10	-0.04	N/A	0.84		
	MLPA	*GABARAPL2* Exon 3	-0.04	0.09	-0.07	-0.14	0.01	0.54	
		GABARAPL2 Intron 3	-0.31	0.21	-0.03	-0.30	0.22	0.97	0.81
TP53	Array-CGH	Zhao et al [52] (exon 17)	0.01	0.52	0.31	0.22			
		CCLE [53]	-0.50	0.16	0.08	N/A	0.95		
	MLPA	*TP53* Exon 4	-0.43	0.24	0.05	0.20	0.84	0.99	
		TP53 Exon 5	-0.41	0.35	0.05	0.06	0.97*	0.96	0.94

*P ≤ 0.05

†For MLPA probes, except for DU145 cell line which is represented by one replicate, average probe ratios obtained from duplicate reaction of cell lines were used.
r : Pearson correlation.